Dynamical mean-field theory and weakly non-linear analysis for the phase separation of active Brownian particles

Thomas Speck,[1] Andreas M. Menzel,[2] Julian Bialké,[2] and Hartmut Löwen[2]

[1]*Institut für Physik, Johannes Gutenberg-Universität Mainz, Staudingerweg 7-9, 55128 Mainz, Germany*
[2]*Institut für Theoretische Physik II, Heinrich-Heine-Universität, D-40225 Düsseldorf, Germany*

Recently, we have derived an effective Cahn-Hilliard equation for the phase separation dynamics of active Brownian particles by performing a weakly non-linear analysis of the effective hydrodynamic equations for density and polarization [Phys. Rev. Lett. **112**, 218304 (2014)]. Here we develop and explore this strategy in more detail and show explicitly how to get to such a large-scale, mean-field description starting from the microscopic dynamics. The effective free energy emerging from this approach has the form of a conventional Ginzburg-Landau function. On the coarsest scale, our results thus agree with the mapping of active phase separation onto that of passive fluids with attractive interactions through a *global* effective free energy (mobility-induced phase transition). Particular attention is paid to the square-gradient term necessary for the dynamics. We finally discuss results from numerical simulations corroborating the analytical results.

I. INTRODUCTION

The separation of a suspension of passive colloidal particles into liquid and vapor is a complex and rather well-studied phenomenon [1]. For phase separation to occur, sufficiently strong attractions between particles have to be present. This is well understood from the perspective of thermodynamics: The potential energy gained by the suspension forming the dense phase compensates for the loss of entropy.

Recently, a separation into dense and dilute regions has also been reported for "active" suspensions composed of self-propelled colloidal particles [2–4]. For a brief perspective on this phenomenon see Ref. 5, and Refs. 6–9 for more general recent reviews on various aspects of active matter. By constantly converting external energy into directed motion, such systems of self-propelled particles can be driven into a non-equilibrium steady state. Steady dynamic states of orientationally ordered collective motion can arise [10–19], but also swirling and turbulent-like situations were observed [20–23]. There are different possibilities how to provide the energy input in experiments: light, if sufficiently strong, can create a temperature gradient leading to self-thermophoresis [24, 25], or lead to the local demixing of a binary solvent [26]. In most experiments, however, energy is provided chemically, *e.g.*, through the decomposition of hydrogen peroxide [27] or the release of stored ions [28]. All of these mechanisms are based on a symmetry breaking on the particle level. In the case of spherical Janus particles [29] it is provided by different surface properties of typically two distinct hemispheres.

What is intriguing from a fundamental perspective is that the phase separation of active colloids strongly resembles the phase separation in a passive suspension, but occurs also for *purely repulsive* particles. This has been demonstrated convincingly in computer simulations of a minimal model by a number of groups [30–33]. This minimal model of active Brownian particles incorporates the two basic physical ingredients: volume exclusion and persistence of motion, *i.e.*, particles interact via repulsive potentials (or hard-core exclusion) and every particle has an orientation along which it "swims" at constant speed. The particle orientations evolve independently and stochastically. The microscopic reason for particle accumulation is simple (see, *e.g.*, the kinetic model in Ref. 31): Due to the persistence of the self-propelled motion and the excluded volume, particles block each other on the time scale that it takes for orientations to decorrelate. If the mean collision time is shorter (the suspension is sufficiently dense) clustering ensues. Simulations with idealized boundary conditions show that hydrodynamic interactions due to the solvent can modify these time scales [34, 35]. On the other hand, experiments [4] indicate that for colloidal swimmers the basic scenario is robust, and we will neglect hydrodynamic interactions in the following.

Even if the microscopic mechanism is known, the collective large-scale and phase behavior is still highly non-trivial. Tailleur and Cates have been the first to realize that the phenomenon of phase separation in active systems can be explained by the dependence of the particle mobility on local density [36–38]. The resulting theory is referred to as "mobility induced phase separation" (MIPS). Within this framework it has been demonstrated that the temporal evolution of the coarse-grained density can be mapped onto that of an effective bulk free energy. Good agreement between particle-resolved simulations of active Brownian particles and the numerical solution of the coarse-grained density has been demonstrated [39, 40]. In analogy to the mean-field treatment of phase separation in passive suspensions (with attractive forces), binodal and spinodal lines are identified from the minima and inflection points of the effective bulk free energy, respectively, from which a schematic phase diagram is constructed. Both lines merge at a single critical point. The resulting free energy involves only the density,

a result that is based on the elimination of the polarization through neglecting temporal and spatial derivatives (a detailed discussion follows in Sec. III). In particular, this treatment produces a bulk free energy that does not contain a term that penalizes sharp interfaces. It is thus not able to describe the dynamics of domain formation and coarsening. In order to cure this shortcoming, Cates and coworkers have argued that the dynamics has to be augmented by gradient terms that are not derivable from a free energy [39, 41].

The purpose of the present paper is to follow the complete path from the microscopic dynamics to the large-scale Cahn-Hilliard equation [42] and to give a comprehensive derivation of the results obtained in Ref. 43. To this end we derive effective hydrodynamic equations and perform a weakly non-linear analysis [44]. Our results refine the MIPS scenario in two ways: (i) Our systematic derivation leads to a square-gradient term that completes the bulk free energy to the conventional Ginzburg-Landau form. (ii) We clarify the validity of a description in terms of the density alone and show that it can strictly hold only close to the loss of linear stability. The physical reason is that this instability occurs on length scales that are larger than the persistence length of the directed motion and thus allow us to study the coarse-grained density alone. The microscopic "trapping" of particles due to their directed motion then enters as an effective attraction. Quenching deeper into the two-phase region, the non-equilibrium nature of active suspensions will become evident again and the coupled evolution of density and orientation has to be considered.

Our paper is organized as follows: In Sec. II, we discuss active Brownian particles as a minimal model for self-propelled disks. Starting from the full many-body dynamics, we sketch the derivation of the effective hydrodynamic equations of motion following Ref. 32. In Sec. III, we explore the consequences of the *adiabatic approximation* of the hydrodynamic equations leading to an effective equilibrium theory for the density alone. We then describe the weakly non-linear analysis in Sec. IV, the ramifications of which are discussed in Sec. V. Conclusions and outlook are given in Sec. VI.

II. MEAN-FIELD THEORY

A. Minimal model for self-propelled disks

The model we study consists of N identical, interacting disks with free diffusion coefficient D_0 moving in a periodic box with area A. Each disk has an orientation $\mathbf{e}_k = (\cos\varphi_k, \sin\varphi_k)^T$ that undergoes free rotational diffusion with diffusion coefficient D_{r}. We consider both diffusion coefficients to be constant and eliminate them through rescaling time $t \mapsto t/D_{\mathrm{r}}$ and length $\mathbf{r} \mapsto \ell\mathbf{r}$, where $\ell \equiv \sqrt{D_0/D_{\mathrm{r}}}$. Neglecting hydrodynamic interac-

tions, the coupled equations of motion become

$$\dot{\mathbf{r}}_k = -\nabla U + v_0\mathbf{e}_k + \boldsymbol{\xi}_k, \tag{1}$$

where $U = \sum_{k<k'} u(|\mathbf{r}_k - \mathbf{r}_{k'}|)$ is the potential energy (in units of the thermal energy) with pair potential $u(r)$, and the noise correlations due to the solvent read

$$\langle \boldsymbol{\xi}_k(t)\boldsymbol{\xi}_{k'}^T(t')\rangle = 2\mathbf{1}\delta_{kk'}\delta(t-t'). \tag{2}$$

The orientational angles obey

$$\langle \dot{\varphi}_k(t)\dot{\varphi}_{k'}(t')\rangle = 2\delta_{kk'}\delta(t-t'). \tag{3}$$

We are thus left with two dimensionless parameters defining a non-equilibrium state point: the number density $\bar{\rho} \equiv N\ell^2/A$ and the free speed v_0.

B. Derivation

An equivalent representation of the equations of motion Eq. (1) is given by the Smoluchowski equation

$$\partial_t\psi_N = \sum_{k=1}^{N}\left\{\nabla_k\cdot[(\nabla_k U) - v_0\mathbf{e}_k + \nabla_k] + \frac{\partial^2}{\partial\varphi_k^2}\right\}\psi_N \tag{4}$$

for the joint probability $\psi_N(\{\mathbf{r}_k,\varphi_k\};t)$ of particle positions and their orientations. This representation is convenient since it allows for systematic approximations. Following Ref. [32], we aim to derive a closed equation of motion for the one-point particle density

$$\psi_1(\mathbf{r}_1,\varphi_1;t) \equiv \int d\mathbf{r}_2\cdots d\mathbf{r}_N \int d\varphi_2\cdots d\varphi_N\, N\psi_N \tag{5}$$

integrating out all other particles. The *local density*

$$\rho(\mathbf{r},t) \equiv \int d\varphi\, \psi_1(\mathbf{r},\varphi;t) \tag{6}$$

corresponds to the probability of finding a particle at position $\mathbf{r}$ at time t.

Inserting Eq. (4) into the time derivative of Eq. (5), we obtain

$$\partial_t\psi_1 = -\nabla\cdot[\mathbf{F} + v_0\mathbf{e}\psi_1 - \nabla\psi_1] + \partial_\varphi^2\psi_1 \tag{7}$$

dropping the particle index. The force $\mathbf{F}(\mathbf{r},\varphi;t)$ is due to interactions of a (fixed) particle with its surrounding particles averaged over their accessible positions and orientations. It thus couples to the two-point density

$$\psi_2(\mathbf{r}_1,\varphi_1,\mathbf{r}_2,\varphi_2;t) \equiv \int d\mathbf{r}_3\cdots d\mathbf{r}_N$$
$$\times \int d\varphi_3\cdots d\varphi_N\, N(N-1)\psi_N. \tag{8}$$

Even in a homogeneous suspension of active particles, the force $\mathbf{F}(\mathbf{r},\varphi;t)$ does not vanish. Rather, there is a force
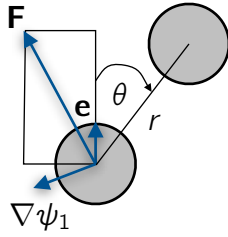
FIG. 1: Sketch of the force decomposition. The total average force $\mathbf{F}$ is approximately decomposed into a component along the particle orientation $\mathbf{e}$ and along the local one-point density gradient $\nabla\psi_1$. Also shown is one of the surrounding particles at distance r enclosing the angle θ with the particle orientation.

imbalance on each particle due to the directional motion, which implies a higher probability to find another particle in front of it (looking along its orientation) than behind. One finds $F \equiv \mathbf{e} \cdot \mathbf{F} = -\rho\zeta\psi_1$ with the force imbalance quantified through the coefficient [32]

$$\zeta \equiv \int_0^\infty \mathrm{d}r \; r[-u'(r)] \int_0^{2\pi} \mathrm{d}\theta \; \cos\theta g(r,\theta) \qquad (9)$$

with pair distribution function $g(r,\theta)$. Here, θ is the angle between the orientation of a particle and the displacement vector to another particle at a distance r, see Fig. 1. For an inhomogeneous density $\rho(\mathbf{r})$ also ζ depends in principle on the position $\mathbf{r}$. Close to the dynamical instability of the homogeneous density profile we can neglect this spatial dependence and in the following we assume $\zeta = \zeta(\bar{\rho}, v_0)$ to be a state function.

We decompose the force

$$\mathbf{F} = (\mathbf{e} \cdot \mathbf{F})\mathbf{e} + \delta\mathbf{F} \approx (\mathbf{e} \cdot \mathbf{F})\mathbf{e} + F_\parallel \nabla\psi_1 \qquad (10)$$

into a component along the particle orientation due to the force imbalance and a component $\delta\mathbf{F}$ perpendicular. The later describes an "evasive" motion leading to an effective diffusion. Hence, as a closure we only keep the projection of $\delta\mathbf{F}$ onto the density gradient $\nabla\psi_1$. This is a good approximation as long as $|\delta\mathbf{F}| \sim 1$ is much smaller compared to $\bar{\rho}\zeta \sim v_0 \gg 1$ [cf. Eq. (41)]. Rearranging $\nabla\psi_1 \cdot \mathbf{F}$ leads to the formal expression[1]

$$F_\parallel \approx \frac{[\nabla\psi_1 - (\mathbf{e} \cdot \nabla\psi_1)\mathbf{e}] \cdot \mathbf{F}}{|\nabla\psi_1|^2}. \qquad (11)$$

Inserting Eq. (10) into Eq. (7), the mean-field evolution equation for the joint probability of position and orientation thus reads

$$\partial_t\psi_1 = -\nabla \cdot [v(\rho)\mathbf{e} - D_\mathrm{e}\nabla]\psi_1 + \partial_\varphi^2\psi_1 \qquad (12)$$

---

[1] There is a typo in Ref. [32] below Eq. (11) (it should read $|\nabla\psi_1|^2$ in the expression for D). Consequently, Eq. (10) is not an expansion in the density gradient but follows from the different magnitudes, $\mathbf{e} \cdot \mathbf{F} \gg F_\parallel$.

with effective diffusion coefficient $D_\mathrm{e} \equiv 1 - F_\parallel > 0$ and effective speed

$$v(\rho) \equiv v_0 - \rho\zeta \qquad (13)$$

depending on the local density.

For a homogeneous suspension, $\rho(\mathbf{r}, t) = \bar{\rho}$ is a constant. Assuming D_e also to be constant, Eq. (12) formally corresponds to the stochastic evolution of a single self-propelled particle. We can then calculate the mean-square displacement and from that the long-time diffusion coefficient [45]

$$D(\bar{\rho}, v_0) = D_\mathrm{e}(\bar{\rho}) + \frac{1}{2}[v(\bar{\rho})]^2. \qquad (14)$$

This relation has indeed been confirmed in several numerical studies of active suspensions [32, 39], and therefore in the remainder we will treat $D_\mathrm{e}(\bar{\rho})$ as spatially uniform and independent of the speed. This constitutes our final approximation closing Eq. (12). Clearly, D_e corresponds to the diffusion coefficient of the passive suspension ($v_0 = 0$).

To briefly conclude, Eq. (12) describes the evolution of the active suspension on a coarse-grained level, into which the effects of microscopic particle interactions enter through two effective parameters: D_e and the force imbalance ζ. Within the theory, every state point $(\bar{\rho}, v_0)$ is fully characterized by these two parameters. However, the theory cannot make predictions about their values, for which we would have to make further assumptions or measure them in particle-resolved computer simulations (as has been done in Ref. 32).

C. Hydrodynamic equations

For the evolution of the local density Eq. (6) one finds

$$\partial_t\rho = -\nabla \cdot [v(\rho)\mathbf{p} - D_\mathrm{e}\nabla\rho] \qquad (15)$$

using Eq. (12). This equation expresses number conservation with a particle current that is given by a diffusive term $-D_\mathrm{e}\nabla\rho$ and a current $v\mathbf{p}$ proportional to the *polarization* or orientational order parameter

$$\mathbf{p}(\mathbf{r}, t) \equiv \int \mathrm{d}\varphi \; \mathbf{e}\psi_1(\mathbf{r}, \varphi; t). \qquad (16)$$

For $\mathbf{p} \neq 0$, particles in a coarse-grained volume have a preferred orientation leading to a net particle current. This orientation evolves according to

$$\partial_t\mathbf{p} = -\nabla P(\rho) + D_\mathrm{e}\nabla^2\mathbf{p} - \mathbf{p} \qquad (17)$$

with pressure

$$P(\rho) \equiv \frac{1}{2}v(\rho)\rho \qquad (18)$$

resulting from the directed motion of the particles. Inserting Eq. (13), for sufficiently large ζ this pressure
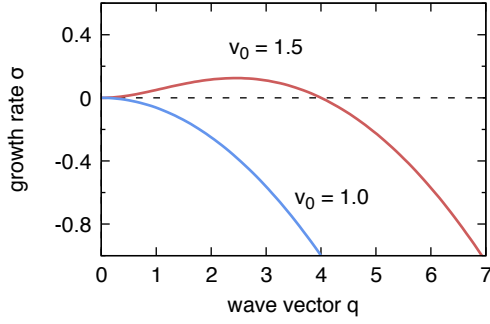
FIG. 2: Dispersion relation $\sigma(q)$ for two speeds v_0 below and above the critical speed $v_c \simeq 1.15$ corresponding to $\bar\rho\zeta = 1$ and $v_* = 1$. For $v_0 > v_c$ large-scale perturbations (*i.e.*, small wave vectors $q > 0$) become unstable.

becomes a non-monotonic function of density. Hence, for local densities $\rho > v_0/(2\zeta)$ the density gradient, $\nabla P(\rho) = P'(\rho)\nabla\rho$, changes sign so that particles migrate towards denser regions. The second term in Eq. (17) is akin to a "viscosity term" and the last term describes the local relaxation due to the rotational diffusion.

While we have derived Eqs. (15) and (17) starting from the full microscopic Smoluchowski equation (4), the formal structure of the result is not surprising since it has to reflect macroscopic conservation laws and symmetries. Note that run-and-tumble dynamics leads to the same hydrodynamic equations (cf. Ref. 37 and see Appendix A). Although colloidal self-propelled particles of course move in a solvent, the decision to neglect hydrodynamic interactions places the resulting effective hydrodynamic theory into the field of what is sometimes called "dry active matter" [7]. Similar equations are, *e.g.*, obtained in the Toner-Tu continuum treatment [11] of polar active systems [46, 47]. In that case the alignment of orientations leads to nonlinear terms of the polarization $\mathbf{p}$ in Eq. (17) and thus to dynamical collective behavior.

D. Linear stability

Clearly, any constant density with $\mathbf{p} = 0$ is a stationary solution of Eqs. (15) and (17). To study the stability of the homogeneous density $\bar\rho$ with respect to small perturbations, we set $\rho = \bar\rho + \delta\rho$ and rewrite the equations of motion (15) and (17)

$$\partial_t\delta\rho = -\alpha\nabla\cdot\mathbf{p} + D_e\nabla^2\delta\rho + \zeta\nabla\cdot(\mathbf{p}\delta\rho), \quad (19)$$

$$\partial_t\mathbf{p} = -\beta\nabla\delta\rho + D_e\nabla^2\mathbf{p} - \mathbf{p} + \zeta\delta\rho\nabla\delta\rho. \quad (20)$$

Here, we have separated the non-linear terms and defined the coefficients

$$\alpha \equiv v(\bar\rho) = v_0 - \bar\rho\zeta, \quad \beta \equiv \frac{1}{2}(v_0 - 2\bar\rho\zeta). \quad (21)$$

Dropping the non-linear terms and inserting solutions of the form $\delta\rho, \mathbf{p} \propto e^{\sigma(q)t + i\mathbf{q}\cdot\mathbf{r}}$, we obtain the dispersion relation

$$\begin{aligned}\sigma(q) &= -\frac{1}{2} - D_e q^2 + \frac{1}{2}\sqrt{1 - 4\alpha\beta q^2} \\ &= -(D_e + \alpha\beta)q^2 - (\alpha\beta)^2 q^4 + \mathcal{O}(q^6),\end{aligned} \quad (22)$$

which quantifies the growth rate of a perturbation with wave vector q (see Fig. 2). Hence, for $D_e + \alpha\beta < 0$ a smooth perturbation of the homogeneous density on small q does not decay anymore but grows, leading to a dynamical instability. Solving this condition implies an *instability line* $v_c(\bar\rho)$ of critical speeds (for simplicity we only consider the smaller solution) such that for $v_0 > v_c$ the suspension becomes linearly unstable. The growth rate $\sigma(q)$ is maximized for the wave length

$$q_0^2 = \frac{1}{4}\left(\frac{1}{\alpha\beta} - \frac{\alpha\beta}{D_e^2}\right), \quad (23)$$

which will thus dominate the morphology at early stages after the onset of the instability.

III. ADIABATIC APPROXIMATION

For slowly varying fields, we neglect the temporal derivative as well as the viscosity term in Eq. (17) to obtain the adiabatic solution

$$\mathbf{p} \approx \mathbf{p}_{ad} = -\nabla P = -\frac{1}{2}\nabla(v\rho). \quad (24)$$

Inserting Eq. (13) for the density-dependent effective speed $v(\rho)$ there are two possibilities to rewrite this expression, which we now discuss.

A. Relaxation of density

First, we push the gradient to the right,

$$\mathbf{p}_{ad} = -\frac{1}{2}(v_0 - 2\rho\zeta)\nabla\rho. \quad (25)$$

We have assumed that ζ is constant, which holds for the homogeneous density profile. Inserting $\mathbf{p}_{ad}$ into Eq. (15) leads to the simple evolution equation for the density

$$\partial_t\rho = \nabla\cdot[\mathcal{D}(\rho)\nabla\rho] \quad (26)$$

with collective diffusion coefficient

$$\mathcal{D}(\rho) = D_e + \frac{1}{2}(v_0 - \rho\zeta)(v_0 - 2\rho\zeta). \quad (27)$$

The global density at which the diffusion coefficient switches its sign is given by the condition $\mathcal{D}(\bar\rho) = 0$. It signals the onset of a dynamical instability at which the homogeneous density profile $\rho(\mathbf{r}, t) = \bar\rho$ becomes (linearly) unstable and small perturbations start to grow. Using the coefficients introduced in Eq. (21) we find $\mathcal{D}(\bar\rho) = D_e + \alpha\beta$ and therefore $\sigma(q) = -\mathcal{D}(\bar\rho)q^2$ demonstrating that it is the same instability discussed in the previous section.

B. Free energy

Alternatively, to illustrate the concept of an effective free energy as advocated by Cates and Tailleur [36, 37], we cast the evolution equation for the density into a form that involves the functional derivative of a potential function, which thus can be interpreted as an effective free energy. This is achieved by pulling out the Nabla operator,

$$-v\mathbf{p}_{\mathrm{ad}} = \frac{1}{2}\nabla\left[v_0^2\rho - \frac{3}{2}v_0\zeta\rho^2 + \frac{2}{3}\zeta^2\rho^3\right]. \tag{28}$$

Inserting this result into Eq. (15), we now find the evolution equation

$$\partial_t\rho = -\nabla\cdot(v\mathbf{p}_{\mathrm{ad}} - D_{\mathrm{e}}\nabla\rho) = \nabla^2\frac{\delta F_{\mathrm{ad}}}{\delta\rho} \tag{29}$$

implying the functional

$$F_{\mathrm{ad}}[\rho] = \int \mathrm{d}^2\mathbf{r}\, f_{\mathrm{ad}}(\rho(\mathbf{r})) \tag{30}$$

with bulk free energy density

$$f_{\mathrm{ad}}(\rho) = \frac{1}{2}\left(D_{\mathrm{e}} + \frac{v_0^2}{2}\right)\rho^2 - \frac{1}{4}v_0\zeta\rho^3 + \frac{1}{12}\zeta^2\rho^4. \tag{31}$$

We will refer to this function as the adiabatic free energy density. It has the typical form of a Landau function often encountered in the study of critical phenomena and phase transitions [48].

In order to discuss the phase diagram following from the adiabatic solution, we rewrite the bulk free energy density as a symmetric function plus a linear term,

$$f_{\mathrm{ad}}(\rho) = f_{\mathrm{ad}}(\rho_0) + \mu(\rho - \rho_0)$$
$$+ \frac{1}{2}\left(D_{\mathrm{e}} - \frac{v_0^2}{16}\right)(\rho - \rho_0)^2 + \frac{1}{12}\zeta^2(\rho - \rho_0)^4, \tag{32}$$

with

$$\rho_0 \equiv \frac{3v_0}{4\zeta}, \qquad \mu \equiv \rho_0\left(D_{\mathrm{e}} + \frac{v_0^2}{8}\right). \tag{33}$$

The bulk free energy density becomes a non-convex function for speeds

$$v_0 \geqslant 4\sqrt{D_{\mathrm{e}}} \equiv v_*, \tag{34}$$

where $v_*(\bar\rho)$ depends on the global density. As the speed is increased beyond v_*, the suspension enters the two-phase region. A common-tangent construction to minimize the bulk free energy density in the non-convex region leads to

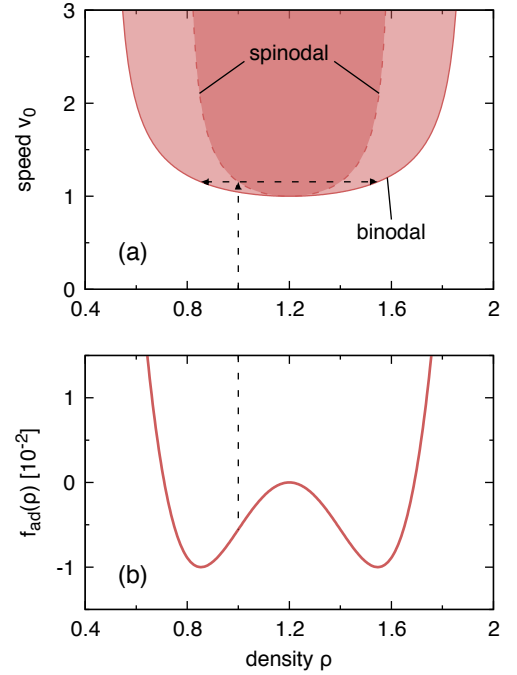$$\left.\frac{\partial f}{\partial \rho}\right|_{\rho_\pm^{(\mathrm{bi})}} = \mu \tag{35}$$

FIG. 3: Schematic phase diagram for the adiabatic mean-field free energy density Eq. (32) using $\zeta(v_0) = (3v_0)/(4\bar\rho_*)$ with critical point at $\bar\rho_* = 1.2$ and $v_* = 1$ (corresponding to constant $D_{\mathrm{e}} = 1/16$). (a) Spinodal (dashed) and binodal (solid) lines in the (ρ, v_0)-plane. Indicated is the behavior for global density $\bar\rho = 1$: Increasing the propulsion speed v_0 (vertical arrow), the homogeneous profile loses linear stability when reaching the spinodal at v_{c}. For any quenched v_0, the coexisting densities are predicted by the points on the binodal (horizontal arrows for v_{c}). (b) The corresponding tilted free energy density $f_{\mathrm{ad}}(\rho) - \mu(\rho - \rho_0)$ for $v_0 = v_{\mathrm{c}}$. The continuation of the dashed line from the upper panel indicates the inflection point for $\bar\rho = 1$.

with the coexisting densities

$$\rho_\pm^{(\mathrm{bi})} = \rho_0 \pm \frac{\sqrt{3}}{4\zeta}\sqrt{v_0^2 - v_*^2}. \tag{36}$$

Consequently, the coefficient μ represents the chemical potential, which according to Eq. (35) is equal for the two coexisting phases of densities $\rho_\pm^{(\mathrm{bi})}$ identifying the *binodal*. Since $f_{\mathrm{ad}}''(\rho) = \mathcal{D}(\rho)$, the inflection points defining the mean-field *spinodal*

$$\rho_\pm^{(\mathrm{sp})} = \rho_0 \pm \frac{1}{4\zeta}\sqrt{v_0^2 - v_*^2} \tag{37}$$

coincide with the limit of linear stability as expected.

The physical picture following from this discussion is thus that of an active suspension undergoing a phase separation into a dense and a dilute phase, see Fig. 3 for an illustration. Although we study a system that is incessantly driven away from thermal equilibrium, its large-scale evolution is apparently that of an effective equilibrium suspension, where the speed v_0 plays the role of

an inverse temperature. In particular, the existence of a free energy-like functional guarantees a unique stationary state in which this functional becomes minimal,

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \int \mathrm{d}^2\mathbf{r}\, \frac{\delta F}{\delta \rho}\frac{\partial \rho}{\partial t} = -\int \mathrm{d}^2\mathbf{r}\, \left|\nabla \frac{\delta F}{\delta \rho}\right|^2 \leqslant 0 \qquad (38)$$

inserting Eq. (29). In qualitative agreement with numerical simulations [30, 31, 39], the mean-field theory predicts a binodal enclosing the two-phase region and a spinodal within this region. The condition

$$\rho_0 = \frac{3v_*}{4\zeta_*} = \bar{\rho}_* \qquad (39)$$

with $\zeta_* \equiv \zeta(\bar{\rho}_*, v_*)$ defines the critical point $(\bar{\rho}_*, v_*)$ at which binodal and spinodal meet.

To conclude this section we comment on two points: First, the functional Eq. (30) does not contain a gradient term penalizing sharp interfaces between low and high density phase, and we will come back to this point in Sec. IV B. Second, the pressure $P(\rho)$ [Eq. (18)] does not have to be equal in the coexisting phases. The interface will consist of particles pointing into the dense phase (otherwise particles will leave the interface back into the dilute phase). Hence, the orientation $\mathbf{p}$ does not vanish in the interface allowing for a jump of $P(\rho)$ following Eq. (24). In addition to the pressure $P(\rho)$, from a free energy density $f(\rho)$ one can derive the "thermodynamic" pressure

$$P_f(\rho) = -\frac{\partial(af)}{\partial a} = -f(\rho) + \rho f'(\rho) \qquad (40)$$

with $P_f \neq P$, where a is a (small) area with locally homogeneous density $\rho \propto 1/a$. This pressure now becomes equal for the coexisting densities Eq. (36).

IV. WEAKLY NON-LINEAR ANALYSIS

A. Dominating mode

Employing the adiabatic solution Eq. (24) allows to reduce the original hydrodynamic equations to a single equation of motion for the density alone, which, moreover, can be cast into a form involving an effective free energy. We now attempt to study the behavior of the hydrodynamic equations in a more systematic way employing a small expansion parameter [43]. To this end we will consider the state points $(\bar{\rho}, v_c)$ along the instability line with force imbalance coefficients

$$\bar{\rho}\zeta_c = \frac{3}{4}v_c - \frac{1}{4}\sqrt{v_c^2 - v_*^2} \qquad (41)$$

following from the condition $D_e + \alpha_c\beta_c = 0$. We study speeds

$$v_0 = v_c(1 + \varepsilon) \qquad (42)$$

close to the instability line with $|\varepsilon| \ll 1$. We expand the coefficients $\alpha(v_0) = \alpha_c + \alpha_1\varepsilon + \mathcal{O}(\varepsilon^2)$ and $\beta(v_0) = \beta_c + \beta_1\varepsilon + \mathcal{O}(\varepsilon^2)$ into Taylor series assuming that they are analytic functions of the speed v_0. To leading order $\alpha\beta \approx -D_e + \sigma_1\varepsilon$ with new coefficient

$$\sigma_1 \equiv \alpha_c\beta_1 + \alpha_1\beta_c. \qquad (43)$$

Hence, the fastest growing wave vector Eq. (23) behaves as $q_0 \sim \sqrt{\varepsilon}$ and the growth rate of structures with this wave vector becomes $\sigma(q_0) \approx -\sigma_1 q_0^2\varepsilon \sim \varepsilon^2$. In the following it will be more convenient to employ non-negative $\varepsilon \geqslant 0$ and let $\sigma_1 \mapsto \pm|\sigma_1|$ so that the sign of σ_1 determines on which side of the instability line we are: for $\sigma_1 > 0$ (small) fluctuations decay while for $\sigma_1 < 0$ the suspension has become unstable.

We now aim to derive an equation of motion for the density fluctuations on the scale of the dominating mode. Since these fluctuations evolve on the length $1/q_0$ and grow with time scale $1/\sigma(q_0)$ we rescale length and time leading to

$$\partial_t \mapsto \varepsilon^2\partial_t, \qquad \nabla \mapsto \sqrt{\varepsilon}\nabla, \qquad (44)$$

which we plug into Eqs. (19) and (20).

B. Close to the critical point

In a first step, we expand the local density and orientation as

$$\rho = \bar{\rho} + \sqrt{\varepsilon}c + \varepsilon c^{(1)} + \varepsilon^{3/2}c^{(3/2)} + \cdots, \qquad (45)$$

$$\mathbf{p} = \varepsilon\mathbf{p}^{(1)} + \varepsilon^{3/2}\mathbf{p}^{(3/2)} + \cdots. \qquad (46)$$

To lowest order the magnitude of density fluctuations is thus $\sim \sqrt{\varepsilon}$, *viz.* the response is $\delta\rho \propto \varepsilon^{1/2}$ as expected close to a critical point with mean-field exponent $\frac{1}{2}$. The expansion form for $\mathbf{p}$ has been chosen to match powers. Plugging all expansions back into the hydrodynamic equations together with Eq. (44), we collect terms of the same order ε. To lowest order, we find $\mathbf{p}^{(1)} = -\beta_c\nabla c$ and therefore

$$0 = (D_e + \alpha_c\beta_c)\nabla^2 c, \qquad (47)$$

which is fulfilled for any perturbation $c(\mathbf{r}, t)$ since the expression in brackets corresponds to the instability condition and thus vanishes.

To next order we find $\mathbf{p}^{(3/2)} = -\beta_c\nabla c^{(1)} + \zeta_c(c\nabla c)$ leading to

$$0 = -2g\nabla \cdot (c\nabla c) = -g\nabla^2 c^2 \qquad (48)$$

with another coefficient

$$g \equiv \frac{1}{2}\zeta_c(\alpha_c + \beta_c) = \frac{1}{4}\zeta_c\sqrt{v_c^2 - v_*^2} \geqslant 0. \qquad (49)$$

For non-vanishing c this condition is fulfilled only for $g = 0$, which implies $v_c = v_*$. Hence, the expansion
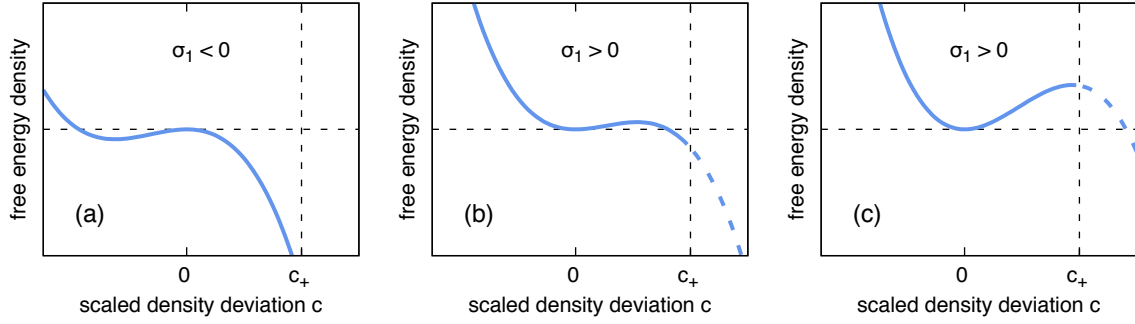
FIG. 4: Possible shapes of the free energy density Eq. (60): (a) In the linearly unstable region for $\sigma_1 < 0$. Due to volume exclusion there is a maximal c_+ (vertical dashed line) at which the density has a minimum. (b,c) For $\sigma_1 > 0$ two cases are possible: (b) the uniform density profile corresponding to $c = 0$ is metastable (the second minimum) or (c) the uniform density profile is globally stable.

Eq. (45) for the density fluctuations is only valid in an ε-environment of the critical point $(\bar{\rho}_*, v_*)$.

Assuming $g = 0$, at the next order we finally obtain an equation of motion

$$\partial_t c = \sigma_1 \nabla^2 c + \zeta_c^2 \nabla \cdot (c^2 \nabla c) - D_e^2 \nabla^4 c = \nabla^2 \frac{\delta F_{1/2}}{\delta c} \quad (50)$$

for the density fluctuations. The right hand side can be expressed as the functional derivative of a potential function (an effective free energy)

$$F_{1/2}[c] = \int d^2\mathbf{r} \left[\frac{D_e^2}{2} |\nabla c|^2 + f_{1/2}(c) \right] \quad (51)$$

with bulk term

$$f_{1/2}(c) = \frac{1}{2} \sigma_1 c^2 + \frac{1}{12} \zeta_c^2 c^4. \quad (52)$$

In contrast to the adiabatic solution Eq. (30), the expansion in ε now includes the customary square-gradient term.

As the final step, we restore the density through inserting $c = (\rho - \bar{\rho})/\sqrt{\varepsilon}$ and reverting the scaling Eq. (44). While the gradient term is invariant, the bulk contribution to the free energy density becomes

$$\varepsilon^2 f_{1/2} = \frac{1}{2} \sigma_1 \varepsilon (\rho - \bar{\rho})^2 + \frac{1}{12} \zeta_c^2 (\rho - \bar{\rho})^4. \quad (53)$$

In order to show that this expression for the bulk is equivalent to Eq. (32) obtained from the adiabatic solution, we first consider

$$D_e - \frac{v_0^2}{16} = D_e - \frac{v_*^2}{16}(1 + \varepsilon)^2 \approx -\frac{v_*^2}{8} \varepsilon \quad (54)$$

up to linear order in ε. On the other hand, we obtain

$$\sigma_1 = -\frac{v_*^2}{8} \quad (55)$$

inserting the force imbalance coefficient Eq. (41) at the critical point $v_c = v_*$. This demonstrates that the adiabatic solution f_{ad} coincides with the result $f_{1/2}$ for the

bulk free energy density of a more systematic expansion close to the critical point. In addition, the later route also yields a square-gradient term that describes the cost of creating density inhomogeneities. This term is important in the coarsening dynamics and allows to predict an interfacial tension from the mean-field theory. Note that such a square-gradient term can already be anticipated from the dispersion relation Eq. (22) from the q^4 term with $(\alpha_c \beta_c)^2 = D_e^2$. The coefficient defines a length D_e that can be interpreted as an effective interaction range.

C. Away from the critical point

We found in the previous section that an expansion in density fluctuations of order $\sim \sqrt{\varepsilon}$ only holds for $v_c = v_*$, i.e., at the critical point. To lift this restriction and to be able to move along the instability line we need to satisfy Eq. (48) while at the same time $g \neq 0$. To achieve this within the ε-expansion of the density one has to demand that the leading contribution is of order $\sim \varepsilon$ (instead of the larger $\sim \sqrt{\varepsilon}$). Effectively, this reduces the magnitude of density fluctuations. We, therefore, arrive at the expansion employed in Ref. [43],

$$\rho = \bar{\rho} + \varepsilon c + \varepsilon^2 c^{(2)} + \cdots, \quad (56)$$

$$\mathbf{p} = \sqrt{\varepsilon}[\varepsilon \mathbf{p}^{(1)} + \varepsilon^2 \mathbf{p}^{(2)} + \cdots]. \quad (57)$$

As before, we collect terms of the same order. At lowest order we again find Eq. (47). The next order already leads to the equation of motion

$$\partial_t c = \sigma_1 \nabla^2 c - 2g \nabla \cdot (c \nabla c) - D_e^2 \nabla^4 c = \nabla^2 \frac{\delta F}{\delta c} \quad (58)$$

for the density fluctuations away from the homogeneous profile, where

$$\mathbf{p}^{(2)} = -\beta_c \nabla c^{(2)} - \beta_1 \nabla c + D_e \nabla^2 \mathbf{p}^{(1)} + \zeta_c c \nabla c \quad (59)$$

has been used. The form of the free energy functional coincides with Eq. (51). In particular, we obtain the

same square-gradient term as before. However, the bulk free energy density now reads

$$f(c) = \frac{1}{2}\sigma_1 c^2 - \frac{1}{3}g c^3. \tag{60}$$

Here the previously introduced coefficients σ_1 and g become the coefficients of the square and cubic term, respectively.

The striking feature of Eq. (60) is that a c^4 term stabilizing the dense phase is missing. Nevertheless, due to volume exclusion, a real suspension will not collapse into a single point but reach the predicted phase-separated state as demonstrated in experiments and simulations. This mechanism of volume exclusion works on length scales corresponding to the particle size implying $c < c_+$ with c_+ corresponding to the maximal density. The damping thus arises from couplings to scales that are not included in the systematic expansion at the above order. Thus the reason for the missing c^4 term is the scale separation between the particle size and the length over which density fluctuations are coarse-grained. The different possible shapes of the free energy density $f(c)$ depending on the system parameters are sketched in Fig. 4. Even without the c^4 term, we can derive a differential equation that describes the instability line, see Appendix B.

V. DISCUSSION

A. Off-critical quenches

The two local free energy densities $f_{1/2}(c)$ and $f(c)$ for the two different scalings can be derived from the same global free energy. To understand this and the role of the asymmetric c^3 term in the free energy density Eq. (60) it is instructive to recall the situation (for passive suspensions) with a symmetric free energy density $f(\varphi) = \frac{a}{2}\varphi^2 + \frac{\kappa}{4}\varphi^4$ with order parameter $\varphi \propto \rho - \bar{\rho}_*$. The Cahn-Hilliard equation models the evolution after a quench from a stable homogeneous density $\varphi = -c_0$ into the two-phase region, where $c_0 = 0$ implies a quench through the critical point. For an off-critical quench, we set $\varphi = c - c_0$ with c the deviations away from the initial homogenous density as before. The Cahn-Hilliard equation then reads

$$\partial_t c = \nabla^2[(a + 3\kappa c_0^2)c - 3\kappa c_0 c^2 + \kappa c^3] - D_e^2 \nabla^4 c \tag{61}$$

dropping constant terms which vanish due to the spatial derivative. Cleary, this result becomes Eq. (58) with $g = 3\kappa c_0$ and $\sigma_1 = a + g^2/(3\kappa)$ when adding a repulsive term $\frac{1}{4}\kappa c^4$ to $f(c)$. The coefficient $a(v_0)$ should be a monotonically decreasing function of speed v_0. It changes sign at $v_0 = v_*$ and becomes negative for $v_0 > v_*$. A given global density $\bar{\rho}$ determines $v_c = v_c(\bar{\rho})$ and $g = g(v_c)$ [Eq. (49)]. Exactly on the spinodal $\sigma_1 = 0$ holds, which implies $a = -g^2/(3\kappa)$ for the quadratic coefficient of the global free energy. Note that at $v_c = v_*$ we have $g = 0$

and, therefore, we recover the function $f_{1/2}(c)$ [Eq. (52)] if in addition we set $\kappa(v_*) = \frac{1}{3}\zeta_*^2$. Hence, $f(c)$ extends the lower order solution $f_{1/2}(c)$ to higher speeds v_0 but with undetermined coefficient $\kappa(v_0)$, which does not follow from the ε-expansion.

B. Nucleation behavior close to the spinodal

The Cahn-Hilliard equation has been derived originally to describe *spinodal decomposition*, the homogeneous, barrierless onset of phase separation throughout the system in response to a quench beyond the spinodal bounding the instable region. Interestingly, for the spinodal itself the Cahn-Hilliard equation predicts a change from a continuous to a discontinuous transition (see, *e.g.* Ref. 49).

To discuss this effect in the present context of active Brownian particles, we symmetrize and rescale the free energy density Eq. (60) (including the repulsive c^4 term) leading to the scaling functions

$$\begin{aligned}\tilde{f}_\pm(\eta) &\equiv \frac{\kappa}{|\sigma_1|^2}[f(c) - f(c_0)] \\ &= \tilde{\mu}_\pm \eta + \frac{1}{2}(\pm 1 - \Gamma)\eta^2 + \frac{1}{4}\eta^4\end{aligned} \tag{62}$$

with scaled density $\eta \equiv \sqrt{\kappa/|\sigma_1|}(c - c_0)$, where $c_0 = g/(3\kappa)$ as in the previous section. These functions depend on the single parameter

$$\Gamma \equiv \frac{g c_0}{|\sigma_1|} = \frac{g^2}{3\kappa|\sigma_1|} \geqslant 0 \tag{63}$$

combining the three coefficients σ_1, g, and κ, with irrelevant $\tilde{\mu}_\pm(\Gamma)$ akin to a chemical potential. Corresponding to the sign of σ_1, $\tilde{f}_-$ describes the effective free energy density in the unstable and $\tilde{f}_+$ in the linearly stable region.

The parameter Γ is defined along the spinodal. In particular, the critical point $(\bar{\rho}_*, v_*)$ corresponds to $\Gamma = 0$ and going away from this point along the spinodal, the value of Γ increases. In Fig. 5 the resulting "phase diagram" is shown in the (η, Γ) plane. Due to the scaling employed in the previous section, the functions $\tilde{f}_\pm(\eta)$ correspond to fixed (small) ε and we now need to distinguish the two cases shown in Fig. 5(a) and (b) depending on which side of the spinodal the system resides. Shown are the binodals calculated from the common tangent construction and the spinodals corresponding to the inflection points, cf. Sec. III B. The uniform density corresponds to setting $c = 0$, *i.e.*, $\eta = -\sqrt{\Gamma/3}$. For $\sigma_1 > 0$ the uniform density is linearly stable but becomes metastable for $\Gamma > \frac{3}{2}$ (it crosses the binodal). This implies that for $\Gamma < \frac{3}{2}$ crossing the spinodal (jumping from $\sigma_1 > 0$ to $\sigma_1 < 0$) corresponds to a continuous transition. In contrast, for $\Gamma > \frac{3}{2}$ the transition becomes discontinuous.

It is instructive to also consider the bifurcation diagrams shown in Fig. 5(c,d). Here we show schematically
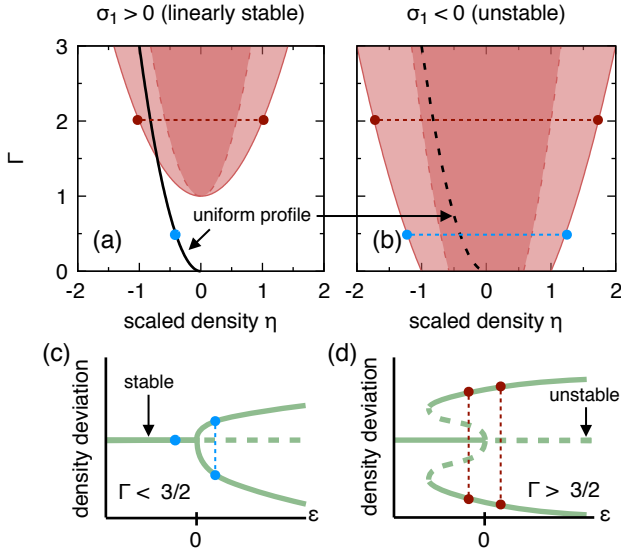
FIG. 5: Schematic "phase diagrams" obtained from the scaling functions (a) $\tilde{f}_+$ and (b) $\tilde{f}_-$, see Eq. (62). Shown are co-existing densities (binodals, outer solid lines) and the limits of linear stability (spinodals, inner dashed lines). For $\sigma_1 > 0$ the uniform density (black line) is linearly stable (it becomes metastable for $\Gamma > \frac{3}{2}$) and for $\sigma_1 < 0$ it is always unstable. (c,d) Sketch of the bifurcation diagrams showing the deviation from the uniform density (the "amplitude", solid lines are linearly stable and dashed lines unstable) vs. the control parameter ε for fixed Γ. The diagrams of (a) and (b) correspond to $\pm|\varepsilon|$ on both sides of the bifurcation point as indicated by the connected dots for two representative values of Γ: (c) Supercritical pitchfork bifurcation for $\Gamma < \frac{3}{2}$ corresponding to a continuous transition. (d) Subcritical bifurcation for $\Gamma > \frac{3}{2}$ corresponding to a discontinuous transition.

the solutions of the amplitude equation, which correspond to the extrema of the effective free energy. Plotted is the density with respect to the homogeneous density as a function of $\varepsilon = v_0/v_c - 1$ for fixed Γ. For any $|\varepsilon| > 0$ close to the instability at $\varepsilon = 0$, the (scaled) solutions can be directly read off Fig. 5(a) for $\varepsilon < 0$ ($\sigma_1 > 0$) and Fig. 5(b) for $\varepsilon > 0$ ($\sigma_1 < 0$) as indicated for two values of Γ. These correspond to a continuous and discontinuous transition, respectively.

C. Non-local speed

Within the weakly non-linear analysis an integrable square-gradient term stabilizing domains appears naturally. To study phase separation kinetics, Cates and coworkers have followed a different route and posit that the active Brownian particles sample the density on a length scale λ larger than the interparticle spacing [39, 41]. To lowest order the speed then becomes a non-local function $v(\hat{\rho})$ with $\hat{\rho} = \rho + \lambda^2 \nabla^2 \rho$ such that

$$v(\hat{\rho}) \approx v(\rho) + v'(\rho)\lambda^2 \nabla^2 \rho. \tag{64}$$

Plugging such a non-local speed into the adiabatic solution Eq. (24) produces the desired square-gradient term but, strikingly, would lead to an equation of motion for the density that also involves non-integrable terms, i.e., it is not longer representable as the functional derivative of an effective free energy. Apparent consequences are discussed in Ref. 41.

Replacing the speed Eq. (13) with the expression Eq. (64), we can again study the systematic ε-expansion of the hydrodynamic equations (15) and (17) close to the instability line. While the lowest order of the ε-expansion remains unchanged, the non-local speed modifies the expression Eq. (59)

$$\mathbf{p}^{(2)} \mapsto \mathbf{p}^{(2)} + \frac{1}{2}\bar{\rho}\zeta_c\lambda^2\nabla^2(\nabla c) \tag{65}$$

for the orientation, where we have used $v'(\rho) = -\zeta$. Consequently, the bulk free energy density Eq. (60) remains the same and the only effect is that the coefficient of the square-gradient term is replaced by

$$D_e^2 \mapsto D_e^2 + \frac{1}{2}\alpha_c\bar{\rho}\zeta_c\lambda^2. \tag{66}$$

Since the additional term is positive, this corresponds to a larger interaction range as one would intuitively expect.

In summary, non-integrable terms were introduced when employing the adiabatic approximation. They do not appear in a systematic treatment. Instead, the systematic treatment naturally produces the well-known squared density gradient term in the effective free energy functional. As a consequence, close to the instability line and at onset (mean-field) phase separation kinetics of active suspensions is predicted to not qualitatively differ from that of passive suspensions even for a non-local speed.

D. Numerical simulations

Although there are already quite a few numerical studies of two-dimensional active Brownian particles [30–33, 43, 51], the determination of the full phase diagram is still an open task. As shown, the mapping to a free energy strictly holds only close to the instability line. Moreover, mean-field treatments are known to yield incorrect quantitative predictions close to critical points due to a diverging correlation length and the ensuing large fluctuations. These fluctuations make numerical sampling notoriously difficult in the vicinity of a continuous transition. This problem is already present in passive suspensions but appears to be even more severe in active suspensions due to two reasons: (i) fluctuations are even larger and (ii) advanced sampling techniques for systems out of equilibrium are not (yet) available.

We examine the data from Refs. [31, 43, 50] obtained from particle-resolved simulations of discoid self-propelled disks with diameter a interacting via the purely
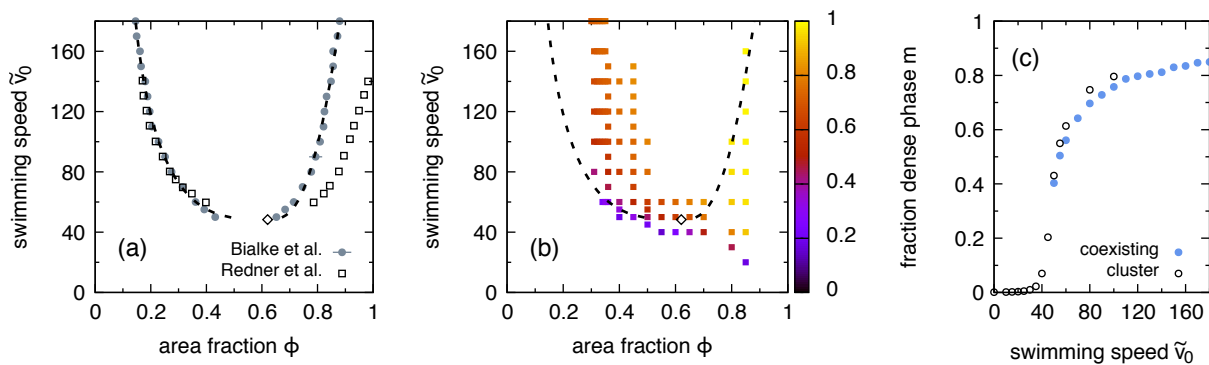
FIG. 6: (a) Coexisting densities for a model suspension, see text for details. Shown are (●) the values from Bialké *et al.* [50] and (□) Redner *et al.* [31]. The dashed lines are fits of Eq. (68) with (◇) indicating the extrapolated position of the critical point. (b) Fitted binodal and critical point as in (a) together with the average fraction m of particles in the largest domain (scale bar on the right, from Ref. 43). Only state points with $m \geqslant 0.1$ are shown. (c) Comparison of m at $\phi = 0.5$ for two methods: (○) from a cluster analysis to identify dense domains and (●) from the coexisting densities via Eq. (69).

repulsive Weeks-Chandler-Andersen potential [52]

$$u(r) = \begin{cases} 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] + \epsilon & (r < 2^{1/6}\sigma) \\ 0 & (r \geqslant 2^{1/6}\sigma) \end{cases} \quad (67)$$

with potential strength ϵ. We set $\sigma = 2^{-1/6}(a/\ell)$ so that repulsive interactions are only present for overlapping particles. In the following, we will present numerical results using the global area fraction $\phi \equiv (a/\ell)^2 \pi \bar{\rho}/4$, which is simply proportional to densities. In principle, the rotational diffusion coefficient D_r is a free parameter of the model. However, for self-propelled colloidal particles it is reasonable to assume (see, *e.g.*, supplementary materials of Ref. 4) that the no-slip boundary condition still holds, which implies $D_r = 3D_0/a^2$. This additional assumption relates the two lengths $a = \sqrt{3}\ell$. For better comparison, in the following we report the original numerical values for the dimensionless speed $\tilde{v}_0 = \sqrt{3}v_0$.

The accurate determination of coexisting densities in numerical simulations is subtle. Redner *et al.* have extracted these densities from histograms [31]. In Ref. 50 a different strategy is followed, in which a non-square simulation box with periodic boundaries is employed. In such a geometry a "slab" of the dense phase forms, which aligns with the shorter box edge. This allows a very accurate determination of density profiles and the coexisting densities. The result for $N = 4,900$ disks at global area fraction $\phi = 0.5$ using $\epsilon = 100k_BT$ is shown in Fig. 6(a). We have obtained data for $N = 10,000$, which agree with the coexisting densities for the smaller system demonstrating that there is no finite-size dependence. Also shown is the data from Ref. 31, who have used $\epsilon = k_BT$ and $N = 15,000$. The low-density branches agree very well while for the high densities the interaction strength plays a more important role. Clearly, the system with the higher ϵ is less compressible. Still, both studies agree qualitatively. Note that below $\tilde{v}_0 \lesssim 60$ fluctuations are so strong that a reliable determination of coexisting densities is not possible anymore.

We have fitted the coexisting densities (the binodal) of Ref. 50 with the functional form

$$\frac{\phi_\pm}{\phi_*} = 1 + a_\pm \left(\frac{v_0}{v_*} - 1\right) + b_\pm \left(\frac{v_0}{v_*} - 1\right)^{1/3} \quad (68)$$

inspired by the famous Guggenheim plot [53]. Here, $a_\pm$ and $b_\pm$ are fit parameters, and $\phi_\pm$ are the coexisting area fractions. This fit works surprisingly well over a wide range of speeds and allows to extrapolate the position of the putative critical point: $\phi_* \simeq 0.62$ and $\tilde{v}_* \simeq 48$. In Fig. 6(b), the fitted coexisting densities are overlaid by results from instantaneous quenching: The passive suspension ($v_0 = 0$) is equilibrated at a given global density. The suspension is then quenched to the final speed v_0 and relaxed to the steady state. A cluster analysis is performed for the recorded configurations to identify dense domains of N_c particles, from which the averaged fraction $m \equiv \langle N_c \rangle / N$ of particles in the largest cluster (*i.e.*, dense domain) is extracted. Note that only state points with $m \geqslant 0.1$ are shown, *i.e.*, the largest domain contains at least 10% of all particles.

Fig. 6(b) is compatible with the proposed mapping to passive phase separation. Below area fraction $\phi \simeq 0.3$ no spontaneous phase separation is observed, which thus roughly indicates the location of the spinodal. There is a larger region around the putative critical point where spontaneous phase separation seems to occur even below the extrapolated binodal, which, however, could be interpreted as a finite-size effect. In a finite system, large fluctuations appear as dense domains but would not lead to full phase separation in a larger system. However, this does not explain the observed phase separation at higher area fraction $\phi \geqslant 0.7$. At these high densities details of the interparticle interactions and the associated particle length scale may become important, which is not captured by the coarse-grained point of view of the weakly nonlinear analysis.

Previously in Ref. [32] we have reported data that would suggest a continuous transition (for $\phi = 0.4$ and $\phi = 0.5$). In line with this observation, in Ref. [43] an abrupt onset of hysteresis for area fraction $\phi < 0.32$ has been observed, which thus agrees with the mean-field prediction of a change from continuous to discontinuous (see Sec. V B). Whether this observed change is a numerical artefact due to the vicinity to a single critical point (as in passive phase separation) or a genuine non-equilibrium effect remains to be investigated.

For completeness, we note that for a single dense domain (as observed at intermediate system sizes) a simple relation between m and the coexisting densities exists via the lever rule,

$$m(v_0) \approx \frac{1/\phi_- - 1/\phi}{1/\phi_- - 1/\phi_+}. \tag{69}$$

This is demonstrated in Fig. 6(c). The result for m obtained from the cluster analysis is slightly larger, which may be attributed to the interfacial particles being counted towards the dense phase.

VI. CONCLUSIONS

To summarize, we have followed a systematic route from the microscopic dynamics to the large-scale Cahn-Hilliard equation to address the phase behavior of active Brownian particles. In Ref. [32], starting from the microscopic dynamics, effective hydrodynamic equations have been derived for the temporal evolution of (weakly perturbed) density and average orientation. Two parameters enter these equations: the free speed v_0 and the global density $\bar{\rho}$. All details of the particle interactions are contained in the single function $\zeta = \zeta(\bar{\rho}, v_0)$, which quantifies the force imbalance due to the interplay between repulsive forces and directed motion.

The hydrodynamic equations exhibit a dynamical instability at $v_c = v_c(\bar{\rho})$, which can be determined through a linear stability analysis [30, 32, 33]. To go beyond the linear regime, we have performed a weakly non-linear analysis [44] close to the instability line using as the small expansion parameter $\varepsilon = v_0/v_c - 1$. This approach yields an equation of motion for the density alone, which is formally equivalent to a Cahn-Hilliard equation. It involves a local effective free energy. We have discussed two expansions holding for different ranges of the speed v_c, which together agree with the scenario of mapping active to passive phase separation *close to the loss of linear stability*. What happens further away from the instability line? Including higher orders of the expansion is of course possible but will lead to a large number of additional terms. Already at the next order the non-equilibrium nature of the active suspension will become manifest since a description in terms of density fluctuations alone is not possible anymore and the time evolution of the polarization has to be included. Hence, there seems to be no real advantage in going to higher orders

since one can also study the original hydrodynamic equations.

One should stress that we have described a mean-field scenario for the one-point density. Stationary two-point correlations are an input to the theory and higher order correlations are neglected. Gaussian noise could be added, which accounts for uncorrelated fluctuations but of course does not restore the missing correlations. For passive suspensions it is well known that mean-field free energies have to be regularized by the Maxwell construction to be thermodynamically valid. Moreover, experimentally and in simulations no sharp loss of linear stability is observed and the "spinodal" line is a pure mean-field concept [54]. Still, one might wonder why for active Brownian particles a mean-field description nevertheless seems to reproduce domain morphologies of particle-based simulations to such a good degree [39]. It thus remains to be tested numerically to which extend the theory outlined here is valid. Some evidence has been discussed in Sec. V D but more detailed numerical investigations are clearly needed.

Despite recent progress there are many open questions even for the simple minimal model studies here. For example, the determination of the full phase diagram is still an open issue both theoretical and numerical. As we have shown here, in the mean-field picture already the instability line has a richer structure than anticipated previously. It will be particularly interesting to study speeds close to v_* in order to determine the nature of the disorder-order transition. However, simulations in this parameter region are hampered by the large critical fluctuations and will require a more detailed study of finite-size effects than available so far. Another open issue is the relation between effective free energy and mechanical pressure [50, 51, 55–57].

Finally, while here we have studied analytically the simplest version of active Brownian particles (namely monodisperse repulsive disks) in two dimensions without alignment and without hydrodynamic interactions, we note that multiple extensions of this model have been discussed in the literature: mixtures of active and passive particles [58], role of attractive interactions [59, 60] and alignment [61], and polydispersity in connection with the glass transition [62, 63].

Appendix A: Run-and-tumble particles

The equivalence of active Brownian particles and run-and-tumble particles on the level of the hydrodynamic equations has been discussed in Ref. 37. For completeness, we briefly sketch the derivation for run-and-tumble particles following the route taken in Sec. II. Again, we consider N identical particles moving in two dimensions. A particle moves with constant speed v_0 along its orientation $\mathbf{e} \equiv (\cos\varphi, \sin\varphi)^T$ (the "run"). After an exponentially distributed random run time with mean τ_r, the particle "tumbles", i.e. it picks a random new orientation. The evolution of the one-point density $\psi_1(\mathbf{r}, \varphi, t)$ is thus given by

$$\partial_t \psi_1 = -\nabla \cdot [\mu_0 \mathbf{F} + v_0 \mathbf{e}\psi_1] - \frac{1}{\tau_r}\psi_1 + \frac{1}{2\pi\tau_r}\rho, \quad \text{(A1)}$$

where $\mathbf{F}$ is the force due to interactions with other particles, μ_0 the bare mobility, and $\rho(\mathbf{r}, t)$ the local number density [Eq. (15)]. The second term describes the "death" of particles with a given orientation φ and the third term their rebirth with uniformly distributed orientation. Dimensionless quantities are introduced through measuring energies in units of ϵ, rescaling time $t \mapsto \tau_r t$, and length $\mathbf{r} \mapsto \ell\mathbf{r}$ with $\ell \equiv \sqrt{\epsilon\mu_0\tau_r}$.

As a closure, we decompose the force $\mathbf{F} \approx -\rho\zeta\mathbf{e}\psi_1 - D_e\nabla\psi_1$, where the first term captures the force imbalance along the orientation slowing down the particles. Since this force is not exactly aligned with the orientation there will also be an "evasive" motion, leading to an effective diffusion on larger scales. This is described through the second term. Inserting the force into Eq. (A1), we thus arrive at [cf. Eq. (12)]

$$\partial_t \psi_1 = -\nabla \cdot [v(\rho)\mathbf{e} - D_e\nabla]\psi_1 - \psi_1 + \frac{\rho}{2\pi} \quad \text{(A2)}$$

with effective speed $v(\rho) \equiv v_0 - \rho\zeta$. It is now straightforward to derive the hydrodynamic equations (15) and (17), which demonstrates the equivalence of active Brownian particles and run-and-tumble particles (at least for weakly perturbed orientations).

Appendix B: Differential equation for instability line

The effective free energy density Eq. (60) is a local function that describes the scaled density deviations in the vicinity of a point on the instability line. Two points along this line have different coefficients σ_1 and g. Still, these cannot be completely independent. Consider a global density ρ_1. Quenching the system to a speed $v_0 = v_c(\rho_1)(1 + \varepsilon)$ past the limit of linear stability, separation into dense and dilute regions will occur. The range of unstable scaled density deviations is $c_-^{(sp)} < c < c_+^{(sp)}$, where $c_-^{(sp)} = \sigma_1/(2g) < 0$ is given by the inflection point of $f(c)$ (the maximal value $c_+^{(sp)}$ is irrelevant for the following argument). Clearly, a suspension with global density $\rho_2 = \rho_1 + \varepsilon c_-^{(sp)}$ should then become unstable beyond the speed $v_0 = v_c(\rho_2)$. We can exploit this consistency condition to derive a differential equation that describes the instability line, see Fig. 7 for a sketch of the derivation.

To this end, consider the derivative

$$\left.\frac{\partial\rho}{\partial v}\right|_{v_1} = \lim_{v_2 \to v_1}\frac{\rho_2 - \rho_1}{v_2 - v_1} = \lim_{v_2 \to v_1}\frac{\varepsilon c_-^{(sp)}}{v_2 - v_1} \quad \text{(B1)}$$

at speed $v_1 = v_c(\rho_1)$. Inserting $\varepsilon = (v_2 - v_1)/v_1$, we obtain

$$\frac{\partial\rho}{\partial v} = \frac{\sigma_1}{2gv}. \quad \text{(B2)}$$

Knowing one point on the instability line, we can integrate this differential equation and invert the solution to obtain the full instability line $v_c(\bar{\rho})$, see Ref. 43 for an example.
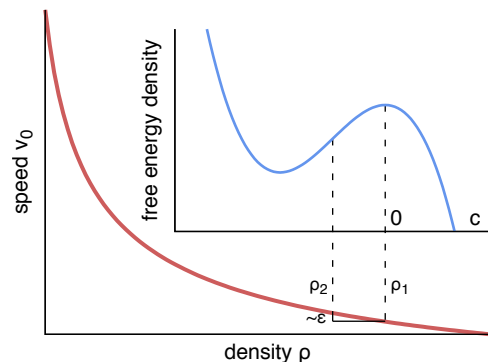

FIG. 7: Sketch for the derivation of Eq. (B2) for the instability line (solid line). For a global density ρ_1 the bulk free energy density $f(c)$ [Eq. (60)] is sketched in the inset, whereby the homogeneous density corresponds to $c = 0$. Densities down to $\rho_2 = \rho_1 + \varepsilon c_-^{(sp)}$ are unstable, which thus corresponds to a second point on the instability line.

[1] V. J. Anderson and H. N. W. Lekkerkerker, Nature **416**, 811 (2002).

[2] I. Theurkauff, C. Cottin-Bizonne, J. Palacci, C. Ybert, and L. Bocquet, Phys. Rev. Lett. **108**, 268303 (2012).

[3] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, Science **339**, 936 (2013).

[4] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck, Phys. Rev. Lett. **110**, 238301

(2013).

[5] J. Bialké, T. Speck, and H. Löwen, J. Non-Cryst. Solids **407**, 367 (2015).

[6] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, Eur. Phys. J. Special Topics **202**, 1 (2012).

[7] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Rev. Mod. Phys. **85**, 1143 (2013).

[8] J. Elgeti, R. G. Winkler, and G. Gompper, arXiv:1412.2692 (2014).

[9] A. M. Menzel, Phys. Rep. **554**, 1 (2015).

[10] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Phys. Rev. Lett. **75**, 1226 (1995).

[11] J. Toner and Y. Tu, Phys. Rev. Lett. **75**, 4326 (1995).

[12] A. A. Evans, T. Ishikawa, T. Yamaguchi, and E. Lauga, Phys. Fluids **23**, 111702 (2011).

[13] A. M. Menzel and T. Ohta, Europhys. Lett. **99**, 58001 (2012).

[14] E. Ferrante, A. E. Turgut, M. Dorigo, and C. Huepe, Phys. Rev. Lett. **111**, 268302 (2013).

[15] A. M. Menzel and H. Löwen, Phys. Rev. Lett. **110**, 055702 (2013).

[16] F. Alarcon and I. Pagonabarraga, J. Mol. Liq. **185**, 56 (2013).

[17] E. Ferrante, A. E. Turgut, M. Dorigo, and C. Huepe, New J. Phys. **15**, 095011 (2013).

[18] C. A. Weber, C. Bock, and E. Frey, Phys. Rev. Lett. **112**, 168301 (2014).

[19] A. M. Menzel, T. Ohta, and H. Löwen, Phys. Rev. E **89**, 022301 (2014).

[20] D. Saintillan and M. J. Shelley, Phys. Fluids **20**, 123304 (2008).

[21] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Löwen, and J. M. Yeomans, Proc. Natl. Acad. Sci. USA **109**, 14308 (2012).

[22] H. H. Wensink and H. Löwen, J. Phys.: Condens. Matter **24**, 464130 (2012).

[23] R. Großmann, P. Romanczuk, M. Bär, and L. Schimansky-Geier, Phys. Rev. Lett. **113**, 258104 (2014).

[24] H.-R. Jiang, N. Yoshinaga, and M. Sano, Phys. Rev. Lett. **105**, 268302 (2010).

[25] A. P. Bregulla, H. Yang, and F. Cichos, ACS Nano **8**, 6542 (2014).

[26] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger, J. Phys.: Cond. Matter **24**, 284129 (2012).

[27] W. F. Paxton, P. T. Baker, T. R. Kline, Y. Wang, T. E. Mallouk, and A. Sen, J. Am. Chem. Soc. **128**, 14881 (2006).

[28] A. Reinmüller, H. J. Schöpe, and T. Palberg, Langmuir **29**, 1738 (2013).

[29] A. Walther and A. H. E. Müller, Soft Matter **4**, 663 (2008).

[30] Y. Fily and M. C. Marchetti, Phys. Rev. Lett. **108**, 235702 (2012).

[31] G. S. Redner, M. F. Hagan, and A. Baskaran, Phys. Rev. Lett. **110**, 055701 (2013).

[32] J. Bialké, H. Löwen, and T. Speck, EPL **103**, 30008 (2013).

[33] Y. Fily, S. Henkes, and M. C. Marchetti, Soft Matter **10**, 2132 (2014).

[34] A. Zöttl and H. Stark, Phys. Rev. Lett. **112**, 118101 (2014).

[35] R. Matas-Navarro, R. Golestanian, T. B. Liverpool, and S. M. Fielding, Phys. Rev. E **90**, 032304 (2014).

[36] J. Tailleur and M. E. Cates, Phys. Rev. Lett. **100**, 218103 (2008).

[37] M. E. Cates and J. Tailleur, EPL **101**, 20010 (2013).

[38] M. E. Cates and J. Tailleur, arXiv:1406.3533 (2014).

[39] J. Stenhammar, A. Tiribocchi, R. J. Allen, D. Marenduzzo, and M. E. Cates, Phys. Rev. Lett. **111**, 145702 (2013).

[40] J. Stenhammar, D. Marenduzzo, R. J. Allen, and M. E. Cates, Soft Matter **10**, 1489 (2014).

[41] R. Wittkowski, A. Tiribocchi, J. Stenhammar, R. J. Allen, D. Marenduzzo, and M. E. Cates, Nat. Comm. **5**, 4351 (2014).

[42] J. W. Cahn and J. E. Hilliard, J. Chem. Phys. **28**, 258 (1958).

[43] T. Speck, J. Bialké, A. M. Menzel, and H. Löwen, Phys. Rev. Lett. **112**, 218304 (2014).

[44] M. C. Cross and P. C. Hohenberg, Rev. Mod. Phys. **65**, 851 (1993).

[45] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian, Phys. Rev. Lett. **99**, 048102 (2007).

[46] E. Bertin, M. Droz, and G. Grégoire, Phys. Rev. E **74**, 022101 (2006).

[47] F. Thüroff, C. A. Weber, and E. Frey, Phys. Rev. Lett. **111**, 190601 (2013).

[48] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, 1995).

[49] A. Novick-Cohen, J. Stat. P **38**, 707 (1985).

[50] J. Bialké, H. Löwen, and T. Speck, arXiv:1412.4601 (2015).

[51] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur, arXiv:1412.5475 (2014).

[52] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).

[53] E. A. Guggenheim, J. Chem. Phys. **13**, 253 (1945).

[54] K. Binder, Philos. Mag. Lett. **87**, 799 (2007).

[55] S. C. Takatori, W. Yan, and J. F. Brady, Phys. Rev. Lett. **113**, 028103 (2014).

[56] F. Ginot, I. Theurkauff, D. Levis, C. Ybert, L. Bocquet, L. Berthier, and C. Cottin-Bizonne, Phys. Rev. X **5**, 011004 (2015).

[57] S. C. Takatori and J. F. Brady, Phys. Rev. E **91**, 032117 (2015).

[58] J. Stenhammar, R. Wittkowski, D. Marenduzzo, and M. E. Cates, Phys. Rev. Lett. **114**, 018301 (2015).

[59] G. S. Redner, A. Baskaran, and M. F. Hagan, Phys. Rev. E **88**, 012305 (2013).

[60] B. M. Mognetti, A. Šarić, S. Angioletti-Uberti, A. Cacciuto, C. Valeriani, and D. Frenkel, Phys. Rev. Lett. **111**, 245702 (2013).

[61] S. K. Das, S. A. Egorov, B. Trefz, P. Virnau, and K. Binder, Phys. Rev. Lett. **112**, 198301 (2014).

[62] R. Ni, M. A. C. Stuart, and M. Dijkstra, Nat. Commun. **4**, 2704 (2013).

[63] L. Berthier, Phys. Rev. Lett. **112**, 220602 (2014).